

Inductive Approach to Predict Personality & Conduct Appraisal Using Advanced Semi-Supervised Multi-Labelled Mining & Learning Technique

¹Shiju George, ²S. Kumaresan

¹PG Scholar, Department of Computer Science and Engineering

²Faculty, Department of Computer Science and Engineering
J.K.K. Nattaraja College of Engineering and Technology, INDIA

Abstract -- In engineering discipline academics are considered at the highest level for individuals but for a successful future, the student must be able to exhibit equally promising personality and conduct. Personality and conduct reveals students ability to cope with the new situations and handle complex problems in long run. In this paper the work is carried out to develop a system that will predict the personality and conduct of a student in advance to know the future results of the engineering students. The system combines supervised and unsupervised learning methods and executes it using multi-labeled data. The personality and conduct details of the students are assigned by the mentor and monitored by HOD of the department for each student in a particular semester. Based on their personality and conduct data that is evaluated at the earlier stage the system executes supervised, unsupervised and semi supervised algorithm to predict new value on the labeled and unlabeled data. Semi-supervised system is an appropriate technique to grade the performance of each student based on the eight traits coined in personality and conduct. Proposed Semi-supervised technique use of multi-labeled data instead of just labeled or unlabeled data to predict the P & C of the student. The predicted value generated by SSLML method over the multi-labeled data is compared with the supervised, unsupervised and semi-supervised technique data using the arm chart for prediction accuracy.

Index Terms- Supervised unsupervised, multi-labeled data, P&C, PCA, Semi-supervised algorithm.

I. INTRODUCTION

We live in a world where vast amounts of data are collected daily. Analyzing such data is an important need. So we need data mining, is an essential process where intelligent methods are applied in order to extract data patterns. In this proposed system, we predict the personality and conduct values of students based on their earlier performances using semi-supervised algorithm with multi-

labeled data. Currently we have got only prediction based on their academic values which can likely to be less helpful in judging a student for a particular responsibility. Unlabeled data are not considered for classification and prediction.

There are eight traits included for the personality and conduct assessment. They are classified under two category one is the personality consisting of Leadership & Team Skills, Discipline & Perseverance, Communication skills, Innovativeness & Creativity and the second is the conduct of a students like Integrity -Courtesy & Kindness, Social Responsibility, Ethical Values, Extra-curricular skills. With respect to these traits the mentor allots marks at the end of the each semester for a mentee. Using these data we predict the personality and conduct values of a particular student in the next semester. Based on the evaluation, marks are awarded for each personality traits and a graph is generated. Here we are able to predict the personality of each student that helps the respective persons to easily evaluate his/her students. Here we are using semi-supervised multi-labeled data. From the earlier analysis it is proved that semi-supervised algorithm is better when compared to supervised and unsupervised algorithms. Also to be more accurate we use multi-labeled data with the semi-supervised algorithm.

II. SURVEY

The surveyed problem is tackled by using the methodology referred to as *semi-supervised learning*. Semi-supervised learning is a special form of classification. Semi-supervised learning addresses the problem by using large amount of unlabeled data, together with the labeled data, to build better classifiers. Because semi-supervised learning requires less human effort and gives higher accuracy, it is of great interest both in theory and in practice. The semi-supervised multi-labeled technique can be used to predict

students' future learning behavior by creating student model that incorporate various traits under the personality and conduct knowledge. The objective of the proposed system is to combine supervised and unsupervised learning methods and automate the system that predicts the individual's traits value by means of the semi-supervision technique. Semi-supervision prediction technique shows promise in developing domain models, such as connecting procedures or facts with the specific sequence and amount of practice items that best teach the methods, and forecasting and understanding student academic, personality and conduct,

also to *predict* future outcomes. Because labeled data is scarce, semi-supervised learning methods make strong model assumptions. Ideally one should use a method whose assumptions fit the problem structure it provides a chance and method to improve the score by giving effective guidelines and description about each trait by the mentors to the students.

Finally it coins the best results by making comparison of the proposed technique with the existing mining & prediction technique. It became important for not only to develop the Personality and conduct appraisal system but also a means to predict the behavior of each students in terms of their position under each traits. These traits are designed that to capture all the requirements necessary for the growth of students in every aspects. So far no such system is built for the assessment of students in engineering field that evaluates the performance of a student and improve their results in personality and conduct based on behavioral modules. The problem statement deals with assessment and appraisal of each student in the various traits. Lack of the P & C system will lead to underrate the students by hiding the qualities in them. There are even techniques that provide only academic performance predication.

III. PROPOSED SYSTEM

The proposed system has got many advantages such as it reduces the risk of improper guidance to the students by lack of information related to P & C. Here we consider the marks to be present in a knowledge base [KB] from where we retrieve the values to predict the personality and conduct values of each student. In the PCA system the mentor analysis each student and assigns marks to them. Calculate Semester Personality & Conduct Point Average (SPCPA). Find total WATS at the end of the final semester. Calculate Course Personality & Conduct average point. Finally map CPCAP in to the graph. The labeled and unlabeled data is clustered by semi-supervised method and further P & C values are predicted for each student by imposing the semi-supervised technique on multi-labeled data.

IV. PCA MODEL

PCA stands for Personality and Conduct Assessment System. The proposed solution to the prediction uses semi-supervised technique which considers both labeled and unlabeled data as an input. It will be implemented by using the high-level language on specific platform.

- Implement the Personality & Conduct Assessment application system
- Applying the existing mining & learning techniques to cluster the unlabeled data in to label one.

- Devise a semi-supervised learning method for multi-labeled data to predict value against each traits and the overall performance of individuals.
- Compared all the techniques with the new one.

ISSUES TO DEAL

- Most of the existing system is based on the Academic Criteria only.
- Designing of various mining algorithm are proposed but the results are not promising.
- Lack of proper integration between the mining and learning process.
- Strategic decision-making is supported but not efficiently.
- Limits ability to engage in process reengineering
- Labeled data are considered at the most for the input. Unlabeled data are not considered for classification and prediction.
- Traits that defines the Personality and conduct of an individual is not addressed which is the key component for the growth of a student.

V. ARCHITECTURE DIAGRAM

The mentors are divided in the ratio 1:20 for the class of strength equal to 60. The mentors got sufficient privileges to interact with PCA system. PCA module consists of PCA sheet for the assessment of the students. The mentor entered grade against each trait for the individual students under his observation. Admin like (Manager/Principal/Dean/HOD) got enough privilege to access the semi-supervised prediction module in addition to the PCA system. The semi-supervised algorithm gets executed over the PCA module to infer individual and overall grades of the students. By this predication method the authority can also monitor the genuine involvement of the mentor in the mentoring process. Each time a prediction is done over a trait by the algorithm the values get stored in the Knowledgebase (KB). The inferred value stored in the KB can be used again to find

new information (facts) for further research over the system. This leads to a comparison among the existing system which uses only the labeled data and the semi-supervised system which uses both labeled and unlabeled. Further the system uses the semi-supervised technique with multi-labeled data. The system will be compared with the arm chart generated for the inputs for the P & C of all students. Results will demonstrate the advantages over the classified system.

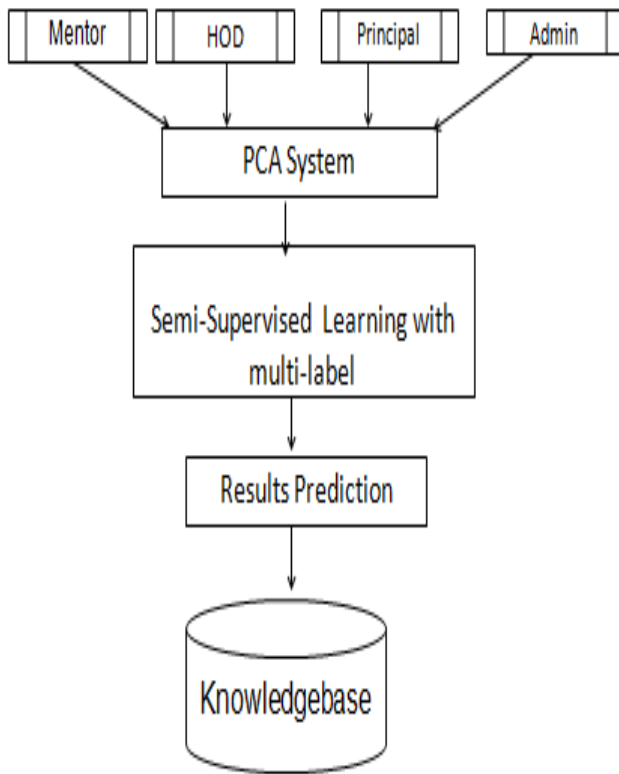


Fig 1. Components of the Architecture

The following modules are created using high level software tools.

a) Personality and Conduct Appraisal System

- (i) User privileges are set (Mentor, HOD, Principal, Admin etc..)
- (ii) Eight traits for each student can be accessed in the P & C Appraisal sheet.
- (iii) Grading will be done by the Mentor

b) Implementation of Semi-Supervised Technique.

- (i) Advanced Single & Multi-Labeled Semi-supervised technique will be implemented
- (ii) The above technique will deal with the unlabeled data and try to define a new cluster so that to map the unlabeled data value to the labeled data to extract predictable information.
- (iii) *Semi-supervised learning* is to use both labeled and unlabelled data together to categorize input values. We can use unlabelled data to build clusters and the

few labeled data points to decide the clusters.

- c) Prediction of Student grades and comparison with original value.
 - (i) Based on the Knowledgebase or labeled values available, the semi-supervised technique will infer expected future grade value of P & C against each trait for the student at the end of the semester.
 - (ii) The original value entered by the mentors at the end of the semester will be compared with the predicted value to ensure correctness.
- d) Make use of existing Data Mining tools to compare the results.
 - (i) Data mining tools are available in the market to verify the results obtained from the proposed technique used.
- e) Regular updating of Knowledgebase (KB).
 - (i) The compared result will be kept in the knowledgebase for the future prediction.
 - (ii) On account of new query with unlabeled data, the KB is used to infer for prediction.

VI. PCA APPLICATION SYSTEM DEVELOPMENT

A) Identify the traits

- Describing each trait.
- Key Quality Indicators
- Observations made against each traits
- Semester Personality & Conduct Point Average (SPCPA) is computed
- $SPCPA (\text{Semester } j) = (G1j + G2j + G3j + G4j + G5j + G6j + G7j + 8j)/8$
- Weighted Average Trait Score (WATS) at the end of the final semester.
- Course Personality & Conduct Point Average (CPCPA)

- Based on CPCPA we will find the quality points statements for each traits.
- Give access rights to mentors and higher authority

B) Personality & Conduct Appraisal Assessment Procedure

This is a continuous system of personality and conduct evaluation through a student counseling system where 20 students are mentored by a faculty mentor, each semester.

C) Grading Procedure

- At the end of each semester (end of S2 for first year B.Tech), the performance of each student is appraised by the faculty mentor on a 5 point scale.
- The eight different personality and conduct traits are awarded letter Grades S(5), A(4), B(3), C(2), D(1) with scores as given in the bracket against each trait.
- The Semester Personality & Conduct Point Average (SPCPA) is computed using the formula $SPCPA (\text{Semester } j) = (G1j + G2j + G3j + G4j + G5j + G6j + G7j + G8j)/8$ which yields a numerical score between 1 and 5 where Gij denote the numerical score corresponding to the trait i for semester j .
- Weighted Average Trait Score (WATS) at the end of the final semester for individual traits are calculated as the weighted average of the scores Gij are scores for trait i obtained in each semester.
- At the end of the eighth Semester, Course Personality and Conduct Point Average CPCPA is calculated as the average of the WATS score for the eight traits.
- Based on the value of CPCPA obtained, a personality and conduct descriptor (Excellent, Very Good, Good or Satisfactory) is assigned to the students.
- The personality and conduct assessment sheet will be provided to the students at the end of the course along with the certificates.
- The records will be kept in the KB of the system.

PERSONALITY & CONDUCT ASSESSMENT												
Student Name :						Department :						
Batch :						University Registration Number :						
Semester weightage for WATS Score		S1&S2	S3	S4	S5	S6	S7	S8	WATS			
Traits	Description	15%	10%	10%	15%	15%	15%	20%	SCORE			
1. Leadership & Team Skills	Responsibility, Motivating others, Negotiation skills, Confidence, Organizational skills, Collaboration & Team spirit.	G12	G13	G14	G15	G16	G17	G18				
2. Discipline & Perseverance	Punctuality, Diligence, Time Management, Money Management, Neatness.	G22	G23	G24	G25	G26	G27	G28				
3. Communication skills	Writing skills, Public speaking, Presentation skills, Persuasiveness, Listening skills.	G32	G33	G34	G35	G36	G37	G38				
4. Innovativeness & Creativity	Native intelligence, Initiative, Entrepreneurship, Multi-disciplinary skills, Ability to think out of the box.	G42	G43	G44	G45	G46	G47	G48				
5. Integrity, Courtesy & Kindness	Truthfulness, Respect for authority, Obedience, Politeness, Consideration, Good manners, Gentleness	G52	G53	G54	G55	G56	G57	G58				
6. Social Responsibility	Social awareness, Fellow feeling, Sense of justice, Voluntary social work, Environmental consciousness.	G62	G63	G64	G65	G66	G67	G68				
7. Ethical Values	Reverence for God, Faith, Modesty, Moral excellence	G72	G73	G74	G75	G76	G77	G78				
8. Extra Curricular Skills	Sports, Games, Performing Arts, Literary skills, Health Consciousness	G82	G83	G84	G85	G86	G87	G88				
SPCPA in each Semester												
Course P & C Point Average (CPCPA)												

CPCPA value	4.5 < CPCPA <= 5	3.5 < CPCPA <= 4.5	2.5 < CPCPA <= 3.5	1 <= CPCPA <= 2.5
Classification	Excellent	Very Good	Good	Satisfactory

Fig. 2 Personality & Conduct Appraisal Assessment Sheet

VII. SEMI-SUPERVISED LEARNING WITH MULTI-LABELED DATA

The solution proposed for the prediction of unknown value is Semi-supervised learning which is a class of supervised learning tasks and techniques that also make use of unlabeled data for training - typically a small amount of labeled data with a large amount of unlabeled data. Semi-supervised learning falls between unsupervised learning (without any labeled training data) and supervised learning (with completely labeled training data).

A. Multi-label classification

There are two main methods for tackling the multi-label classification problem. They are problem transformation methods and algorithm adaptation methods. Problem transformation methods transform the multi-label problem into a set of binary classification problems, which can then be handled using single-class classifiers. Algorithm adaptation methods adapt the algorithm to directly perform multi-label classification. Some of the algorithms used for multi-labeled classification are:-

- K-nearest neighbors: the ML-kNN algorithm extends the k-NN classifier to multi-label data.
- Decision trees: "Clare" is an adapted C4.5 algorithm for multi-label classification; the modification involves the entropy calculations.

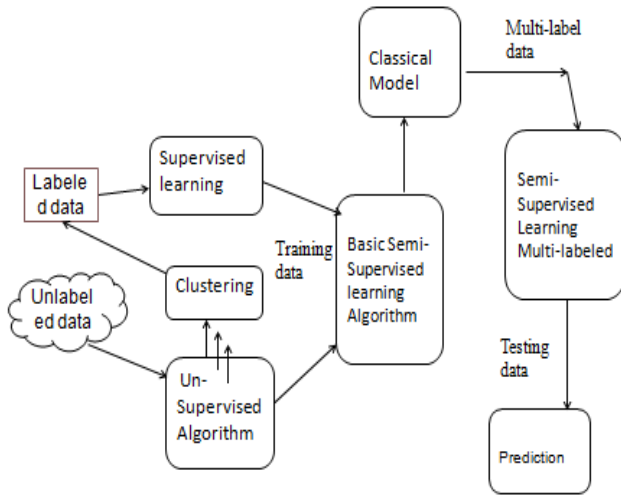


Fig 3. Semi-supervised Learning with Multi-labeled data modeling

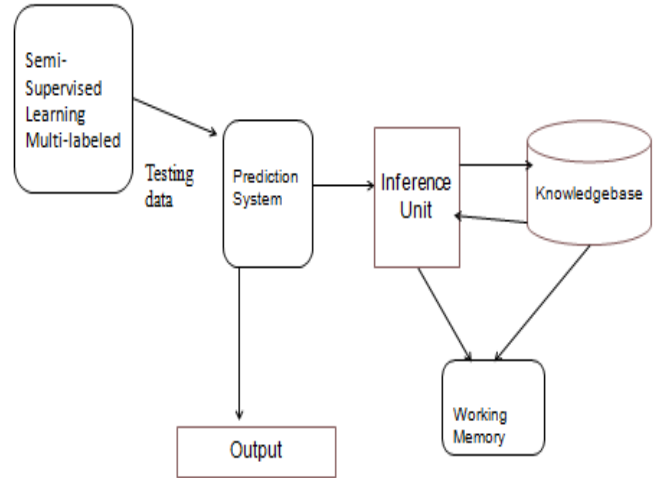


Fig 4. Personality & Conduct Prediction

VIII. PREDICTION METHOD

- Proposed algorithm exhibits the combine features of data mining supervised and unsupervised techniques like classification and clustering algorithm.
- The algorithm inherits the results based on the semi-supervised technique.
- All the attributes or traits value will act as the test-data for the system and will be inputted on the training data-sets.
- Algorithm will get the input in the form of test data and get executed with optimized result. Labeled and unlabeled data are defined
- The results will compare with the independent technologies used in the existing system.

IX. SIMULATION RESULT AND DISCUSSION

Learning method develops the Knowledgebase by classifying and mapping the unlabeled data in to labeled data. Facts in the KB can be utilized for the general overall performance prediction. Performance of each student based on their personality and conduct is evaluated. Based on their grades scored for the P & C traits performance Dynamic evaluation chart can be created and qualities will be awarded. With this information we are able to predict the future performance of the student. Mentors get sufficient feedback to monitor the individuals and give better guidance to improve their trait values in rest of their semesters. Platform to compare the proposed technique with the existing one. Able to find the best semi-supervised algorithm & technique for prediction. Personality & Conduct are presented in more generalized way. Biased decision will be reduced. An organization can see the overall performance of the student's community. The best mining method can be found applied by comparison.

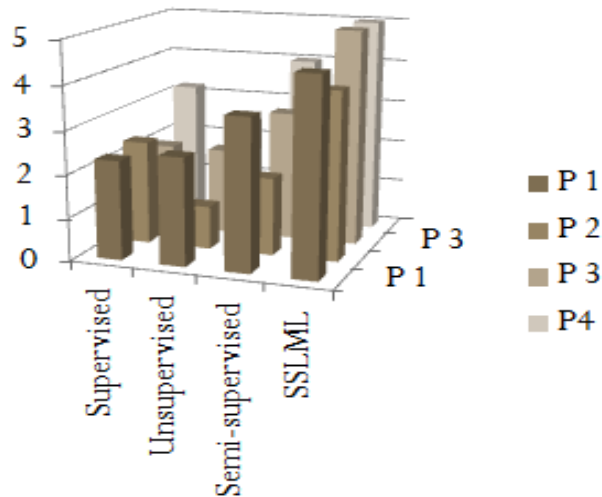


Fig.5 A view of Performance by each method

X. CONCLUSION

Unlabeled data is comparatively easy to gather. Semi-supervised learning can be used to classify the unlabeled data and also it can be used to develop better classifiers. Proposed system can cluster the unclassified traits and its value with similar values together and dissimilar in a independent cluster. The results will be promising with the semi-supervised technique. This paper focuses on predicting student P & C by using advantage of semi-supervised learning with multi-labeled data. Semi-supervised learning method implementation that allows taking advantage of the strengths of supervised and semi-supervised technique. Implement SSLML proposed method that is able to label the unlabeled grades automatically in each iteration and update current classifier to converge it towards similar training of supervised learning. Prediction of Student grades and comparison with original value. Make use of existing Data Mining tools to compare the results. Regular updating of Knowledgebase (KB). The compared result will be kept in the knowledgebase for the future prediction.

XI. REFERENCES

- [1] S. Goldman And Y. Zhou, "Enhancing Supervised Learning With Unlabeled Data," In Proceedings Of The 17th International Conference On Machine Learning, 2000, Pp. 327-334.
- [2] Taruna, S. ; Pandey, M. , "An Empirical Analysis Of Classification Techniques For Predicting Academic Performance "Advance Computing Conference (Iacc), 2014 Ieee International, Publication Year: 2014 , Pp. 523 – 528
- [3] Kcvcv Fgf Javadi,A; Gray,A; Anderson,D ; Berisha V,Semi-Supervised Hierarchy Learning using Multiple-Labeled Data,Machine Learning For Signal Processing,2011 Ieee International Workshop,Publication,Year: 2011,Pages
- [4] Edin Osmanbegovic, M. S. (2012, May). Data Mining Approach For Predicting Student Performance. Journal Of Economics And Business, 10(1), 3-12.
- [5] B. Sen, E. Uçar And D. Delen, "Predicting And analyzing Secondary Education Placement-Test Scores: A Data Mining Approach", Expert System With Application, Volume 39, Issue 10, 2012.
- [6] B.K.Bhardwaj And S.Paul , "Mining Educational Data To Analyze Students Performance", International Journal Advanced Computer Science And Application Vol. 2 No. 6 , 2011 .
- [7] B. M. Bidgoli, D.Koshy, G.Kortemeyer, W.F.Punch, "Predicting Student Performance: An Applicant Of Data Mining Methods With An Educational Web Based System" , 3rd Asee/ Ieee .Frontiers In Education Conference 20004.
- [8] C. Romero And S. Ventura, "Educational Data Mining: A Survey From 1995 To 2005," Expert Systems With Applications, No. 33, Pp. 135–146, 2007.
- [9] C.Romero ans S, Ventura, " Educational Data Mining: A Review of the State of the Art", IEEE Transaction on Systems, Man, and Cybernatics, Vol.40, No.6, 2010.
- [10] D.Kabakchieva, "Predicting Student Performance by using Data Mining methods for classification." ,Cybernetics and Information Technologies, Volume 13,2013.
- [11] E. Osmanbegovic and M. Suljic, " Data mining Approach for Prediction of Student Performance" Economic Review - Journal of Economics & Business Vol. 10, issue 1, 2012.
- [12] I.H. Witten and E. Frank, Data Mining: Practical machine learning tools and techniques. San Francisco:Morgan Kaufmann, 2 ed., 2005.
- [13] I.H. M. Paris, L.S. Affecndy and N.Musthafa, "Improving Performance Prediction using Voting technique in data Mining", World Academy of Science, Engineering and Technology World Academy of Science, Engineering and Technology, Vol 38,2010
- [14] □. Z. Erdogan and M. Timor, "A Data Mining Application in a Student Database," in Journal of Aeronautics and Space Technologies, 2005, vol.2, No. 2, 2005, pp. 53-57.
- [15] [2] K. B. Brijesh and P. Saurabh, "Mining Educational Data to Analyze Students Performance," In International Journal of Advanced Computer Science and Applications, vol. 2, No. 6, 2011, pp. 63-69.
- [16] N. Delavari, S. Phon-Amnuaisuk, and M. R. Beikzadeh, "Data Mining Application in Higher Learning Institutions," In Informatics in Education, vol. 7, No. 1, 2008, pp. 31-54.